

H. VARGA GYULA

AZ IGEKÖTŐS ALAKULATOK FORMAI VIZSGÁLATA

Analízis és szintézis

Néhány éve tekintélyes összefoglalás jelent meg a számítógépes nyelvészetről. Ebben az igekötőkről a következőket olvashatjuk: „Egybeírásnál a szótárbeli elemekkel való összevetés nem jelent problémát. Különírásnál azonban nem tudunk olyan eljárást, mely egy mondatbeli potenciális igekötő (például *oda*, *be* stb.) és egy ige együvé tartozását megerősítené vagy elvetné. Ezért az egyetlen elfogadható eljárásnak a felhasználó megkérdezését tartjuk” (Prószéki 1989, 534 – kiemelés tőlem, V. Gy.). Igaza van a szerzőnek: az elváló és becéklődő igekötős szerkezetek explicit leírása és algoritmikus elemzése még mindig várat magára. Pedig e részprobléma tisztázására az elmélet (nyelvelméleti leírás) és a gyakorlat (gépi fordítás, lemmatizálás stb.) szempontjából egyaránt szükség van.

Az automatikus szintézis oldaláról a magyar igekötős ígét eddig csupán Melcsuk vizsgálta (1958, in Papp 1964, 147–214). Ő abból indul ki, hogy a szövegben igekötő nélküli (az igekötőtől elválasztott) ige is előfordulhat. Ezek összekapcsolására ún. keresési szabályokat állított fel. A dichotomikus osztás elvére épülő szabályok feltételekből és utasításokból állnak. Kiindulási egysége a „két pont közötti szövegrész” (i. m. 161). Azon belül történik a szavak keresése – jobbra és balra. Ha valamelyik szót nem tudja azonosítani (nincs meg a „tőtárban”), akkor meg kell nézni, nincs-e körülötte igekötő. Ennek keresésében az igekötő tipikus szórendi helyeit veszi figyelembe, és logikus lépésekben kapcsolja össze őket. (Vö. i. m. 164–5.) Melcsuknak nem csupán a igekötős ige szétváló elemeinek az összekapcsolása a célja: ő a mondatot vizsgálja, azt akarja más nyelvre fordítani. Az említett reprodukció nála csak részletkérdés. Talán ezért is nem fordított gondot az összekapcsolás finomságaira.

Nem világos a tagmondatok elhatárolása. Márpedig ennek figyelembe vétele nélkül a keresésben a szomszédos tagmondatok elemei is összekapcsolódhatnak. (Pl. a *Ha jön, meg fogod tudni* mondatban a program fordított szórendű igekötős ígét talál – *megjön* –, ti. a vessző nem csak tagmondatokat választhat el.) Esetenként – a szófaj-azonosítás előtt – nem ígéhez is kerestet

igekötőt. (L. a II. 0.00 keresési szabályt, i. m. 164.) Később, a 2. fázisban minden igehez kerestet igekötőt (11. szabály, i. m. 166), ez hosszadalmas és fölösleges. Az igevekről nem esik szó, s a beékelődéseket sem veszi figyelembe.³

Az itt bemutatandó eljárás alapvetően különbözik a Melcsukétól: nem az igehez keresünk igekötőt, hanem a igekötőhöz igét vagy igeveget. Ige ugyanis gyakran előfordulhat igekötő nélkül, de fordítva csak igen-igen ritkán. (Egy fő derivátumait könnyű összeszedni a konkordanciából, de a prefixumos, pl. igekötős származékait szinte lehetetlen.)

A igekötő (PF⁴) formai elemzését két helyen kell elvégezni: a) a szóban (szövegszóban), ha prefixumként szerepel igéjével együtt (egybeírva); b) a szövegben, ha önállóan áll. Az elsőre az analízis fázisában azért van szükség, hogy hozzáférhetővé tegyük az igét (vagy annak valamely származékát) a további formai vizsgálatok számára; a másodikra a szintézisben kerül sor, segítségével lehetővé válik az eredeti alakok reprodukálása. De ez a művelet további eljárások kiindulásául is szolgálhat, pl. lemmatizálás, lexémátár számítógépes készítése konkordanciából, magyarról történő (emberi és gépi) fordítás.

Az automatikus elemzés a következő fázisokban történik:

1. Előkészítés
 - 1.0 Bevitel → szöveg
 - 1.1. Elhatárolás → mondat
 - 1.2. Tagolás → szövegszó
2. Azonosítás → igekötő
3. Elemzés
 - 3.1 Szintetizálás
 - a) Keresés → V/NV
 - b) Összekapcsolás → PF+V/NV
 - c) Korrekció → mondat
 - 3.2. Analízis
 - a) Bontás → PFV/NV
 - b) Ellenőrzés → V/NV
 - c) Korrekció → mondat

³ Pontosabban azt nézi, van-e az igétől „legfeljebb 2 szóval jobbra vagy 3 szóval balra” igekötő (i. h.).

⁴ A továbbiakban használatos rövidítések: V = ige; PF = igekötő; PFV = igekötős ige; NV = igenév, N = főnév; INF = főnévi igenév; PART = melléknévi igenév; ADV = határozói igenév.

1. Előkészítő műveletek

1.0. A nyelv bármely elemét vizsgáljuk is – az elemzés természetétől függetlenül –, az elsődlegesen a szövegben férhető hozzá. Tehát a s z ö - v e g b ől mint közvetlenül adott, egységes egészből kell kiindulnunk. Most, amikor az a feladatunk, hogy a mondat (a kommunikáció) céljainak megfelelő műveleteket végrehajtsuk az igekötős alakulatokon, a bemeneten tagolatlan jelsort (szöveget) kapunk.

1.1. Az első lépés a m o n d a t o k e l h a t á r o l á s a. A mondat-egészek fölismerése egyszerű: a mondatzáró írásjeltől (első mondat esetén szövegkezdettől) mondatzáró írásjelig tartó szövegdarabot mondategységnek minősítjük. A mondategység (tagmondat) elhatárolása már nehezebb feladat. Ez olyan formális-szintaktikai elemzést kívánna, amely messze túlmutat a dolgozat témakörén. Kitűzött célunkhoz erre szerencsére nincs is szükségünk.

Ha ugyanis a magyar mondatokat szemügyre vesszük, azt tapasztaljuk, hogy – függetlenül a tagmondathatártól, az igekötő szórendjétől s attól, hogy az alkotó tagok mennyire távolodtak el egymástól – az igekötő és az ige(név) között n e m á l l h a t s e m m i l y e n í r á s j e l. Ez az apróság a formai elemzésben igen fontos szerepet kap. Ennek alapján határolhatjuk el egymástól a k v á z i - m o n d a t o k a t, melyeken belül az igekötők s a hozzájuk tartozó igék, igenevek keresését, összekapcsolását végezzük. Esetünkben kvázi-mondatnak a két bármilyen írásjel közötti szakaszt (intervallumot) nevezzük.⁵

Meg kell jegyeznünk, hogy az igekötő és az ige(név) között elvileg páros írásjel (gondolatjel, zárójel) is állhat, de ez igen ritkán fordul elő, egyébként kívülről egyszerűen átugorható, belül pedig a szabály szerint kell eljárni. Pl.: *Meg akarja – pontosabban: szeretné – találni.*

1.2. A következő lépés az írásjelekkel elhatárolt egységeknek szöveg-szókra tagolása. Ezt úgy határozhatjuk meg, hogy szóköztől szóközиг terjedő szövegdarab, eljárásunkhoz finomítva: 0 betűtől 0 betűig terjedő karakterlánc (betűsor).

2. Azonosítás

Ebben a fázisban történik az egyes szócsoportok, szövegszók szótári azonosítása és értékelése. Ehhez az igekötők explicit leírására van szükség: vagy

⁵ Lektorai véleményében Sebestyén Árpád hívta fel a figyelmemet a kötőjel szerepére. Az olyan szerkezetekben, mint a *fel- vagy lebontás, át- meg átszövi, meg-megáll* stb. (a Sebestyén Árpád példái) a kötőjel a szótövet, az alapigét helyettesíti, ilyenformán nem sorolandó fel a kvázi-mondatok határát jelző írásjelek között.

tételes felsorolásra, vagy szabály megadására. Esetünkben az előbbi az egyszerűbb. Gyakorlati megfontolásból az igekötők listáján föl kell tüntetnünk az igekötőszerű előtagokat is (pl. *cserben[hagy]*, *tönkre[megy]*, *végre[hajt]*), így programunk majd az összes elváló igét és igenevet kezelni tudja.

Egy PF megkeresése a listán – Melcsuk gondolatát átvéve – „annyit jelent, mint megtalálni azt a maximális hosszúságú betűsort, melyet a keresett szó tartalmaz” (i. m. 161). Vesszük tehát az egyes szövegszókat egészen az írásjelig, és sorban megnézzük, rajta vannak-e a listán. Amennyiben valamelyik szövegszó azonosítása pozitív eredményt ad, vagyis az adott betűsor azonos a lista valamelyik tagjával, akkor – egy másik lista alapján – ellenőrizni kell, nem homonimával van-e dolgunk.⁶ Ha a talált elem valóban PF, akkor ezt regisztráljuk, s ezzel befejeződött az azonosítás fázisa.

Mielőtt tovább elemeznénk a bemenő nyelvi jeleket, tisztáznunk kell egy módszertani kérdést. A most ismertetett lépéssor a szintézist készíti elő. *S z i n t é z i s e n* itt nem szintetikus eljárasmódot értünk, hanem a szövegben analitikusan (széttagoltan) álló igekötős igék, igenevek reprodukálását, a kiinduló alakok helyreállítását. És fordítva, *a n a l í z i s e n* pedig a memóriában tárolt alakból analitikus szerkezet létrehozását, az előképzős lexémák bontását, előkészítését különféle nyelvi műveletek számára. Hangsúlyozzuk: *e l ő* készítését, nem pedig *e l* készítését. Itt és most tehát nem feladatunk a különböző szórendű mondat szerkezetek generálása. Az analízis azonban nemcsak önálló műveletsorként végezhető el, hanem az azonosítási fázisból is elágaztatható. Melcsuk idézett általános szókeresési szabálya megengedi a részleges egybeesést is.

Ha azt írjuk elő, hogy az adott szövegszó (teljes egészében) *e g y e z z e n* meg a lista valamely tagjával, akkor a kimeneten az elváló PF-ek jelennek meg – a szintézishez. Most adjuk meg az azonosítási szabályt úgy, hogy a vizsgált szövegszónak *t a r t a l m a z n i a* kell a lista valamelyik tagját *i s*, mégpedig a szövegszó 1. karakterével kezdődően. Ekkor összegyűjthetjük a szöveg összes PF-et tartalmazó szavát. Ezeket még ebben a fázisban külön kell válogatni, hogy a kimeneten külön jelenjenek meg a *r é s z b e n* azonosítható (tehát PF-et is tartalmazó) és a *t e l j e s e n* azonos jelsorok (PF-ek). Olyan feladatban, amely analízist nem kíván (pl. lemmatizálás), az igekötős szavak közvetlenül feldolgozhatók; más esetben pedig az analízis műveletein kell őket végigfuttatni. A teljes azonossággal

⁶ A homonímia kérdéskörébe itt most nem érdemes belebonyolódni, csupán megjegyzem, hogy az automatikus elemzésnek általában – így ennek is – még mindig ez az Achilles-sarka.

nyert alakok (önálló PF-ek) pedig minden esetben a szintézis irányában továbbítandók.

3. Az elemzés műveletei

3.1. Szintetizálás

A PF megtalálása után meg kell keresnünk annak eredeti (kiindulási) helyét. Keressük tehát azt a PF nélküli V, NV, ritkán N értékű szövegszót, amelynek morfológiai tartozéka a megtalált PF. A kétféle szórend (fordított és megszakított) miatt a keresést két irányban, jobbra és balra kell végezni. A keresésnek két útját járjuk végig: egy nyelvit (félautomatikus) és egy gépit (automatikus).

3.1.1. Félautomatikus szintetizálás

Mint azt máshol már bemutattam, a fordított szórendű szerkezetek jóval nagyobb valószínűséggel (4:1 arányban) fordulnak elő, mint a megszakított szórendűek. Ez azt jelenti, hogy az elváló PF-hez tartozó V vagy NV 75 %-os valószínűséggel közvetlenül a PF előtt áll.⁷ A keresést tehát balra érdemes kezdeni.

A) Kiindulópontunk az írásjel. Ha a keresést balra akarjuk elkezdni, akkor arra kell rákérdeznünk, hogy áll-e a PF előtt írásjel. (Más szóval: mondat elején van-e a PF?) Az igekötő általában a mondat belsejében áll, így a válasz várhatóan nemleges, tehát indulhatunk balra. Ha mégis pozitív választ kapunk, akkor nyilvánvalóvá válik, hogy a szerkezet csakis megszakított szórendű lehet, jobbra meg kell találnunk a V-t vagy NV-t.

A keresett ige – ritkán igenév – szinte mindig *k ö z v e t l e n ü l* az igekötő előtt áll (*nézzük meg, úgy adjuk át* stb.). A fordított szórendű PF-nek ez a tipikus pozíciója. Egy kis részük (kb. 1 %) 1–2 szóval hátrább kerül a V-től, pl.: *mintha puskából lőtték volna ki, nem hajították még el, térjünk most vissza* stb. A PF kivételesen jobban is eltávolodhat a V-től. (Pl.: *Ott folyjon az ifjúi vér ki szívémből*, Petőfi, idézi J. Soltész 1959, 106.) Ez a fajta mondat szerkezet azonban – főleg az írott nyelvi stílusokban – igen ritka.⁸ Így

⁷ Az igekötőtől balra szinte kivétel nélkül V áll (az esetek 98,38%-ában), jobbra pedig többnyire NV (67,41%, ebből 59,51% INF) található (vö. H. Varga 1993).

⁸ J. Soltész Katalin szerint ez „nem ritkaság” (i. h.), azonban a vizsgált közel másfél százezer szövegszó között nem találtam rá példát.

figyelman kívül hagyhatjuk. Elég, ha – Melcsukhoz hasonlóan – a PF előtti 1–2. szót vizsgáljuk.

Jelöljük az egyes szövegszókat az ábécé nagybetűivel, az igekötőt – mint azonosított elemet – PF-fel, kvázi-mondatunk kezdetét és végét pedig ; -vel. Írjunk fel egy tetszőleges, igekötőt is tartalmazó kvázi-mondatot:

; A B C PF D ; (a D itt lehet Ø is).

Fordított szórend esetén tehát vesszük a PF előtti (balra az 1.) szót, ha az V vagy NV – de nem a *volna* segédzó –, akkor megtaláltuk a keresett szót; a PF-et a V elé tesszük, és összekapcsoljuk őket. Az előbbi szimbólumokat használva:

– ha C = ige, akkor a kiírás:

;A	B	PFC	D;
----	---	-----	----

– ha C = *volna*, akkor vesszük a következő (balról a 2.) szövegszót, a B-t;

– ha B = ige, akkor a kiírás:

;A	PFB	C	D;
----	-----	---	----

Amennyiben még az sem ige, akkor nem fordított szórendű a szerkezet: visszatérünk a PF-hez, és a keresést jobbra folytatjuk.

B) Az egyenes megszakított szórendű szerkezetben már nehezebb az eligazodás. Mindenesetre jobbra az 1. szó nem lehet ige, mert akkor – egyenes szórend (helyesebben sorrend, l. alább) lévén – a PF és a V egyetlen szövegszót alkotna.

A szintaktikai elemzésben részletesen taglaltam (vö. H. Varga 1993), milyen elemek és hogyan ékelődhetnek be az igekötős alakulatokba. Tudjuk, hogy igekötős igébe csak IS típusúak ékelődhetnek (*is, se, sem, ne, nem, mégsem, úgysem*, i. m. 317). A keresés és összekapcsolás így történik:

; A B C PF D E;

;A	B	C	D	PFE;
----	---	---	---	------

itt az A, B, C szimbólumok bármilyen szövegszót (Ø-t is) jelölhetnek, D = IS, E = ige.

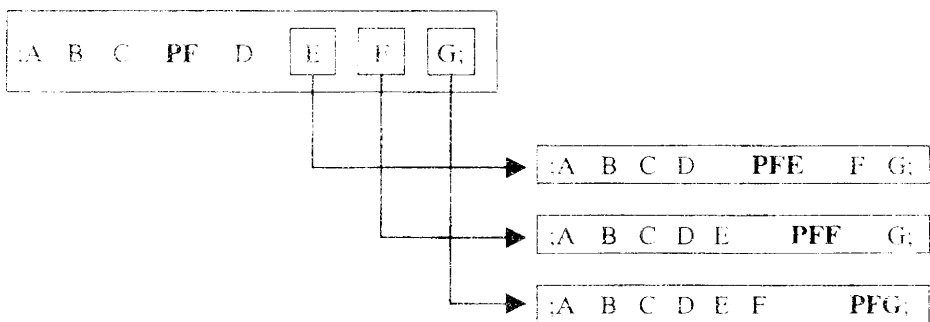
Az igevek esetében másképp alakul a szórend. PART és N ritkán szakad szét, tagjai között ékszóként csak a *nem* állhat. (A hatóképzős alakok a

többi IS típusú ékszót is eltűrik.) Az ADV, főleg pedig az INF alakjaiban igen gyakori és változatos a beékelődés. Azonosításuk egy lista alapján történik. Ennek tartalmaznia kell – a lehetőségek szerint – mindazokat a karakterláncokat, amelyek ékszókként azonosíthatók a PF és a V vagy NV között.

A keresési szabályt általános érvényűvé kell tennünk. A PF utáni első szimbólumról megállapítottuk, hogy ékszónak kell lennie. Ha mégsem az, a keresést nem folytatjuk: a PF-nek minősített elem vagy elliptikus szerkesztésű (igéje az előző mondatban áll), vagy nem is igekötő (téves volt az azonosítás, pl. a homonímia miatt).

A másik lehetőség szerint a PF után valóban ékszó áll. Vesszük a következő (jobbra a 2.) szövegszót. Ha ez V vagy NV, de nem ékszó (pl. *rá sem néz*, de: *rá sem akar nézni*), akkor a szintézis elvégezhető.

A ciklus további részében a következő szimbólummal végzett műveleteket hajtjuk végre mindaddig, amíg a keresett igenevet (itt már igék nem állhatnak!) meg nem találjuk, vagy írásjelig nem érünk.



3.1.2. Automatikus szintézis

Az automatikus elemzést úgy kell elképzelnünk, hogy egy automata bemenetére jelsorok kerülnek. Ezeket egy felismerő egység szegmentálja, pontosabban fölismeri, hogy diszkrét jelek sorozatáról van szó. Ezután a memóriában tárolt listák alapján sorra azonosítja a kapott jelsorok tagjait. Azonosság esetén – annak mértékétől függően – végrehajtja az analízis vagy a szintézis műveletét.

A szintézisnek az előzőekben ismertetett módja nem volt teljesen automatikus, mert a megelőző leírás sem volt tisztán formális. Nem kaptuk meg mindegyik adatunk explicit bemutatását: nem határoztuk meg pl., hogy mi az

az írásjel, ige, igenév stb.⁹ E hiányosságok miatt a rendszer nem lehetett automatikus vezérlésű: kívülről folyton be kellett avatkozni (pl. eldönteni, hogy az adott szó ige-e). Ezt próbáljuk a következőkben kiküszöbölni.

Az *előkészítés* (tagolás) csupán finomítani kell: pontosan meg kell adnunk az írásjeleket, például így: $X_0=, , "$; $X_1=, , , "$; $X_2=, , ; "$; $X_3=, , . "$; $X_4=, , ! "$; $X_5=, , ? "$ stb. Eszerint kvázi-mondatunk definíciója ez lesz: X_{1-5} -ig terjedő jelsor; a szövegszóé pedig ez: X_0 -tól X_{0-5} -ig tartó karaktorsor. A kapott szövegszókat sorszámmal látjuk el, s bármikor hivatkozhatunk rájuk (vagyis azonosíthatjuk őket).

Az azonosítási fázis korábban leírt műveletsorát – az analízis felé vezető lépés nélkül – itt is fölhasználhatjuk. Legföljebb annyi kiegészítést tehetünk, hogy egy tagmondatban (kvázi-mondatban is) csak egy (elvált) PF lehet.

Az igeekötős szóalakok sorrendi (és nem szórendi) típusait a következőképpen vázolhatjuk fel:

1. a) összekapcsolt egyenes sorrend (*fölad*);
b) megszakított egyenes sorrend (*föl is adja*);
2. a) kapcsolt fordított sorrend (*add föl*);
b) megszakított fordított sorrend (*add már föl*).

A négy típus közül három tartalmaz elvált igeekötőt. Az 1. b) jelű szerkezet beékelődése 1–6 elemből áll, ezek meghatározott rendben követik egymást.

PF + IS + KELL + *volna* + MÉG + AZT + *is* + V/NV.

A 2. b) típus szerkezete egyszerűbb:

V + *volna/csak*/MÉG/AZT + *is* + V/NV.

A jelzéseket leegyszerűsítve a három szerkezetet így írhatjuk fel:

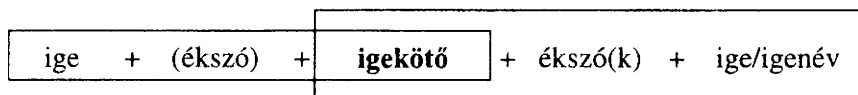
PF + I + K + v + M + A + i + V/NV;

V + Ø + PF;

V + v/c/M/A + i + PF.

Ha a három sorrendi típust egymásra vetítjük, olyan kvázi-szerkezetet kapunk, amely mindenféle igeekötős alakulat modellálására és elemzésére alkalmas.

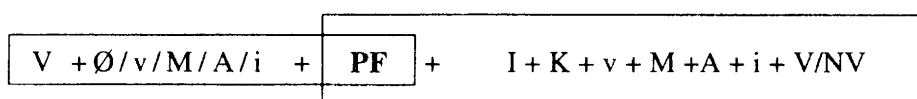
⁹ Az „ige cselekvést, történést stb. jelentő szó”-féle definíciók hiányosságairól vö. Papp 1975, 6.



fordított sorrend

egyeses (megszakított) sorrend

Ugyanez szimbólumokkal jelölve, kiegészítve az egyes ékszótípusokkal ilyen képletet mutat:



A kapott kvázi-szerkezetet szimbólumai lényegében változók, amelyek egy-egy – explicit módon leírható – részhalmaz állandóit reprezentálják. Kvázi-szerkezetünk tehát úgy értelmezhető, mint különböző tengelyek mentén elhelyezkedő részhalmazok rendezett, reprezentatív halmaza. Ez adja meg a *s z i n t a g m a t i k u s* t e n g e l y t, tagjai az igekötős szerkezet elemcsoportjainak reprezentánsai. A *p a r a d i g m a t i k u s* t e n g e l y e k mentén találhatók az igekötős szerkezetek elemcsoportjai, pontosabban azok állandói. E részhalmazoknak természetesen üres (Ø) elemei – így üres metszetei – is lehetnek.

Szintagm. tengely →

	V	v	M	I	PF	I	K	v	M	A	V/NV
Paradigm-	V ₀	v ₀	M ₀	I ₀	–	I ₀	K ₀	v ₀	M ₀	A ₀	V/NV ₀
matikus	V ₁	v ₁	M ₁	I ₁	PF ₁	I ₁	K ₁	v ₁	M ₁	A ₁	V/NV ₁
tengely	V ₂		M ₂		PF ₂	I ₂	K ₂		M ₂	A ₂	V/NV ₂
↓	..		M ₃		...	I ₃	..		M ₃		..
	..		M ₄		...	I ₄	..		M ₄		..
	V _n		M ₅		PF _n	I ₅	K _n		M ₅		V/NV _n

A jelek magyarázata: I = IS, K = KELL / SZABAD / VAN, v = *volna*, M = MÉG, A = AZT, V = ige, V/NV = ige vagy igenév. A számok az adott részhalmaz tagjait, a ₀ indexek a halmaz üres elemét jelölik (pl. I₀ = 0, I₁ = *is*, I₂ = *ne*, I₃ = *nem* stb.).¹⁰

¹⁰ Az egyes típusok részletezését lásd H. Varga 1993, 317.

A) A megszakított szórendű alakulatok szerveződési szabályairól a korábban említett helyen explicit leírást adtam. Ennek fölhasználásával az automatikus keresést biztonságosabb *j o b b r a* kezdenünk, az előző lépések megfordításával.

A szintagmatikus tengely első *a z o n o s í t o t t* tagja a PF, a második az IS részhalmaz valamelyik eleme (ez lehet \emptyset is!). Ellenőrizzük, hogy a PF-et követő szó eleme-e az I részhalmaznak (IS típusú-e?).

1. Jól szerkesztett mondatban PF után kizárólag akkor állhat IS, ha ez a PF megszakított szórendű igekötős alakulat tagja (pl. *meg is kérdezi, kérdezd is meg*, de: **kérdzd meg is*). Megfordítva: ha PF után IS áll, ott kell lennie a PF-hez tartozó V-nek vagy NV-nek is.

a) $PF + IS + V \rightarrow PFV + IS$, pl. *ki sem néz \rightarrow kinéz + sem*. A kimeneten megkaptuk a keresett lexémát, az adatot tároljuk, vagy továbbítjuk, mindenestre a szintézis művelete befejeződött.

b) $PF + IS + KELL$: szintézisük a $PF + KELL$ mintájára végezhető el.

c) $PF + IS + VAN$: a $PF + VAN$ típus szabályai szerint járunk el.

d) A névszói és a melléknévi igenévi megbontott alakulatok helyreállítása az l. a) alatt leírtak mintájára megy végbe:

$PF + nem + N \rightarrow PFN + nem$ (pl.: *meg nem elégedés – meglegedés*);

$PF + nem + PART \rightarrow PFPART + nem$ (pl.: *ki nem mondott – kimondott*);

$PF + IS + PART (IMP) \rightarrow PFPART (IMP) + IS$ (pl.: *el sem képzelhető – elképzelhető*).

2. A legnagyobb nehézséget a KELL típusúak jelentik. Mivel ezek előfordulhatnak kombinált beékelődésben is, igen körültekintően kell eljárunk. Megnézzük, hogy a KELL utáni pozícióban nem a *volna* segédszó, esetleg MÉG vagy AZT típusú ékszó áll-e (pl. *ki tudok nézni, ki tudtam volna nézni, ki (is) tudtam volna már nézni* stb.). Ilyenkor az ékszókat sorra átlépjük: a legelső nem ékszó lesz a keresett elem, tehát az összekapcsolás elvégezhető.

A teljes biztonság kedvéért azonban még egy korrekciós szűrőt beiktathatunk a művelet sor végére. Tudjuk, hogy $PF + (IS) + KELL$ szerkezet után a mondatban – valahol – INF-nek kell állnia. Ellenőrizzük az első nem ékszó („INF-gyanús” elem) végződését: az utolsó két karakternek NI-nek kell lennie.

A KELL típusúak közül a *kell* és a *lehet* segédigeként nem ragozódik (igeként ragozzuk: *Te is kellesz a próbához; Ott lehetek én is*; segédigeként nem: *Feltétlenül el kell jönnöd; Ott lehet maradnom* stb.). Ezek után a személytelen (segéd)igék után az INF ragozódik.¹¹ Az ékszókat tekintve ez

¹¹ A többi személytelenül használatos igénk (*illik, fáj v. jólesik kimondanom* stb., vö. Keresztes 1953, 341) ékszóként nem vagy csak kivételesen használatos.

elvileg a SZABAD típusra is érvényes, közülük azonban csak a *szabad* viselkedik így. A *képes* mellett nem ragozódik az INF, a többi ékszóként is ritka, ragvonzó használata pedig már nyelvhelyességi kérdéseket is érint. (Pl.: *ki szükséges válogatnom > szükséges kiválogatnom; bele muszáj nyugodnotok > muszáj belenyugodnotok* stb.)

A *kell*, *lehet*, *szabad* után az INF végződése a következőképpen alakul:

<i>adni</i>	<i>kérni</i>	<i>kötni</i>
adnom	kérnem	kötnöm
adnod	kérned	kötnöd
adnia	kérnie	kötnie
adnunk	kérnünk	kötnünk
adnotok	kérnetek	kötnötök
adniuk	kérniük	kötniük

A ragok rendszere tehát:

V _{vel.}	V _{ill.}	V _{pal.}	V _{lab.}
-ni	-ni	-ni	
-nom	-nem	-nöm	
-nod	-ned	-nöd	
-nia	-nie	-(nie)	
-nunk	-(nünk) ¹²	-nünk	
-notok	-netek	-nötök	
-niuk	-(niük)	-niük	

Az 1. személyt jelölő *-nom*, *-nem*, *-nöm* kötött alternációjú allomorfok végeredményben a *-NOM* morféma alternánsai (vö. Antal 1964, 11 kk.). Így az INF az alábbi végzésekkel rendelkezhet:

-NI; -NOM, -NOD, -NIA; -NUNK, -NOTOK, -NIUK.

A fent említett három ékszó után – közvetlenül vagy közvetve – álló INF tehát ezekre a karaktercsoportokra végződhet.

3. Hasonló módon ellenőrizhető a VAN típusúak utáni ADV is: az összekapcsolásra minősített szövegszónak *-va*, *-ve*, *-ván*, *-vén* végződésűnek kell lennie.

¹² A zárójeles alakok egybeesnek a másik palatális oszlop megfelelő alakjaival.

Összefoglalva: a megszakított szórendű szerkezetek szintézise úgy történik, hogy a szintagmatikus tengely PF utáni tagjait vesszük egyesével, s ellenőrizzük, hogy azok elemei-e a paradigmatisz tengely valamelyik részhalmazának.

Addig állunk a szintagmatikus tengely adott pontján, amíg az ékszó azonosítása meg nem történt. Ilyenkor a paradigmatisz tengely mentén lépünk egyet-egyet, amíg a részhalmaz elemei el nem fogynak (ez egy ciklus), majd átugrunk a következő részhalmazra, s a ciklus indul előlről. Ha megtaláljuk az ékszót, lépünk egyet a szintagmatikus tengelyen. A művelet általában 2–3 (ritkábban több, legfeljebb 7) lépésből áll¹³, az egyes lépéseken belül pedig 2–3 (legfeljebb 6) ciklus található.¹⁴

A szintézis utolsó lépése, a konkatenáció (összekapcsolás) helyességének biztosítására egy ellenőrző szűrőt alkalmazunk. Ennek az a feladata, hogy az ékszót követő első, a listán nem azonosítható elemről kiderítse, hogy az megfelel-e az általunk keresett NV valamelyikének.

B) Amennyiben a halmazok metszeteinek azonosítása nem járt eredménnyel, a keresést a másik irányban kell folytatnunk. A két írásjel által közrefogott igekötő lemmatizálása viszonylag egyszerű feladat.¹⁵ Vissza kell térni az előző mondathoz, benne olyan szövegszót keresni, amely a talált PF-fel kezdődik. Ha ráakadtunk, a vizsgált (önálló) PF-et töröljük, helyébe írjuk a talált PFV-t.

A keresési fázisok fölcserélésével annyit most biztosan tudunk, hogy az önállóan szereplő PF nem megszakított (hanem fordított) szórendű szerkezet tagja. Az elemzés jóval egyszerűbb, mint az előző fázisban. Itt a művelet többnyire 1, ritkábban 2, de legfeljebb 3 lépésből áll, s 3 rövid ciklust tartalmaz (l. a mondatmodellt).

Ellenőrző szűrőt ide nem ágyazhatunk be, mert a PF előtti V-t formailag nem tudjuk azonosítani: nincs róla formális leírásunk. Itt (V + *volna* + PF), valamint a megszakított szórendű igék esetében (PF + IS + V) el kell fogadnunk, hogy az adott helyzetben csak ige fordulhat elő. Elvileg természetesen itt is van lehetőség az ellenőrzésre. Nyilván a V adatainak explicit kifejtése

¹³ A vízszintes tengelyen ennyiféle ékszó lehet, pontosabban eggyel kevesebb, mert a 7. lépésben az INF-et kell azonosítani.

¹⁴ A paradigmatisz tengelynek ennyi részhalmaza (ékszótípusa) van. L. a mondatmodellt.

¹⁵ Itt tkp. nem is szintetizálásról, hanem analógiás kiegészítésről van szó, hisz a mondat nem tartalmaz igét, pl.: [*Holnap visszaadod?*] – *Vissza!* → *visszaad*.

kell hozzá: vagy a végzések azonosítása (vö. pl. Jánoska kísérletét: 1967, 464–68), vagy más szófaj-meghatározó automata.

C) A szintézis most kifejtett változata már valóban automatikus: a jelentés teljes mellőzésével, tisztán formai, alaktani és szórendi tulajdonságok alapján a szöveg bármely pontján található igekötőhöz kikereshető a vele egy lexémát alkotó V, NV vagy N, majd egy művelettel (konkatenáció) helyreállítható a kiinduló alak.

Annak igazolására, hogy ez a művelet sor valóban automatikusan (külső beavatkozás nélkül) működik, a leírás és a folyamatábrák alapján személyi számítógépre – programozók segítségével – Pascal nyelvben készítettünk egy szintetizáló programot.

A program egy, a billentyűzetről adott mondatban megkeresi az igekötőt s a tőle különálló igét vagy igenevet, és ha egymáshoz tartoznak, összekapcsolja őket. A program az igekötőhöz tartozó igét két írásjel vagy írásjel és mondat eleje között keresi.

Az adatok tárolása és kezelhetősége érdekében a memóriában kialakított listák a következők:

- az A jelű lista az igekötőket tartalmazza;
- a B listán a PF utáni első részhalmaz elemei vannak (IS típusúak);
- a C listán azok a szavak szerepelnek, amelyek második beékelődő elemek (ha nincs IS, akkor elsők); itt találhatók tehát a KELL, a SZABAD és a VAN típusúak;
- D lista: az összes többi ékszó helye;
- az E lista: az igenevek végzéseit sorolja föl.

A program elindítása után a következő tájékoztató szöveg jelenik meg: „A program egy adott mondatban megkeresi az igekötőhöz tartozó igét, majd összekapcsolja őket.” Rövid szünet után ez a felirat olvasható: „Kérem, a mondatot gépelje be, és az ENTER billentyű lenyomásával zárja.”

Ezután be kell írni a mondatot a fenti utasítások szerint. Pl.: *Meg kellett volna még azt is kérdeznem, ki utazik most el.* (Enter) A gép válasza:

Igekötős ige(név):

Meg } kérdeznem
el { utazik.

Új mondat? (I/N)

A jelek magyarázata:

- } = megszakított egyenes sorrend,
- \ = kapcsolt fordított sorrend,
- { = megszakított fordított sorrend.

Ha az I billentyűt nyomjuk meg, akkor új mondatot írhatunk be, ha az N-t, akkor a program STOP állapotba kerül. Mint látható, a kiírásban a sorrendi típus szimbóluma is megjelenik.

Ha a mondat végére nem tettünk írásjelet, a gép figyelmeztet, és újra kéri a mondatot. Amennyiben az „igekötőt” nem tudja mivel összekapcsolni, a (tag)mondatot változatlan formában írja ki. Pl.: *Tollat meg papírt is kell hozni.* (Enter.) A kiírás:

Tollat meg papírt is kell hozni. – Ebben a mondatban nincs igekötő.

I r o d a l o m

Antal László 1964. A formális nyelvi elemzés. Gondolat Kiadó.

Jánoska Sándor 1967. A magyar ige automatikus toldalékolásának egy modellje. NytudÉrt. 58, 464–468.

Keresztes Kálmán 1953. A személyragos főnévi igenév használatáról. Nyr. 77: 340–352.

Melcsuk, Igor 1964. A magyarról oroszra történő gépi fordítás szabályainak ismertetése. (Ford.: S. Varju Anna) In: Papp Ferenc (szerk.): Matematikai nyelvészet és gépi fordítás a Szovjetunióban. OMKDK 6, 147–213.

Papp Ferenc 1975. Konkordancia: írói szótár előkészítése számítógépen. Nyr. 99: 351–355.

Prószéky Gábor 1989. Számítógépes nyelvészet. Számítástechnika-alk. Vállalat, Bp.

J. Soltész Katalin 1959. Az ősi magyar igekötők. Akadémiai Kiadó, Bp.

H. Varga Gyula 1993. A magyar igekötők szórendi típusai. MNy. LXXXIX, 312–318.